

Wed Jul 18 08:51:37 2001

us-09-457-066-1_copy_829_1188.rst

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 2001, 00:04:02 ; Search time 1176.35 Seconds
(without alignments)
2892.870 Million cell updates/sec

Title: US-09-457-066-1_COPY_829_1188
Perfect score: 360
Sequence: 1 gcttttgttttgaagaad.....gcagaggagcacaggagga 360

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues
Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

, Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: gb_estl1: *
2: gb_est2: *
3: gb_est3: *
4: gb_est4: *
5: gb_est5: *
6: gb_est6: *
7: gb_est7: *
8: gb_est8: *
9: gb_est9: *
10: gb_est10: *
11: gb_est11: *
12: gb_est12: *
13: gb_est13: *
14: gb_est14: *
15: gb_est15: *
16: gb_est16: *
17: gb_est17: *
18: gb_est18: *
19: gb_est19: *
20: gb_est20: *
21: gb_est21: *
22: gb_est22: *
23: gb_est23: *
24: gb_est24: *
25: gb_est33: *
26: gb_est34: *
27: gb_est35: *
28: gb_est36: *
29: gb_est37: *
30: gb_est38: *
31: gb_est39: *
32: gb_est40: *
33: em_estba: *
34: em_estfun: *
35: em_esthum1: *
36: em_esthum2: *
37: em_esthum3: *
38: em_esthum4: *
39: em_esthum5: *
40: em_esthum6: *
41: em_esthum7: *
42: em_esthum8: *
43: em_esthum9: *

44: em_esthum10: *
45: em_esthum11: *
46: em_esthum12: *
47: em_esthum13: *
48: em_esthum14: *
49: em_esthum15: *
50: em_esthum16: *
51: em_esthum17: *
52: em_esthum18: *
53: em_esthum19: *
54: em_esthum20: *
55: em_esthum21: *
56: em_esthum22: *
57: em_esthum23: *
58: em_esthum24: *
59: em_esthum25: *
60: em_esthum26: *
61: em_esthum27: *
62: em_esthum28: *
63: em_estin1: *
64: em_estin2: *
65: em_estin3: *
66: em_estin4: *
67: em_estin5: *
68: em_estom1: *
69: em_estom2: *
70: em_estov1: *
71: em_estov2: *
72: em_estpl1: *
73: em_estpl2: *
74: em_estpl3: *
75: em_estpl4: *
76: em_estpl5: *
77: em_estpl6: *
78: em_estpl7: *
79: em_estpl8: *
80: em_estpl9: *
81: em_estpl10: *
82: em_estro1: *
83: em_estro2: *
84: em_estro3: *
85: em_estro4: *
86: em_estro5: *
87: em_estro6: *
88: em_estro7: *
89: em_estro8: *
90: em_estro9: *
91: em_estro10: *
92: em_estro11: *
93: em_estro12: *
94: em_estro13: *
95: em_estro14: *
96: em_estro15: *
97: em_estro16: *
98: em_estro17: *
99: em_estro18: *
100: em_estro19: *
101: em_estro20: *
102: gb_est25: *
103: gb_est26: *
104: gb_est27: *
105: gb_est28: *
106: gb_est29: *
107: gb_est30: *
108: gb_est31: *
109: gb_est32: *
110: gb_est41: *
111: gb_est42: *
112: gb_est43: *
113: gb_est44: *
114: gb_est45: *
115: gb_est46: *
116: gb_est47: *

190:	gb_estill0:*
191:	gb_estill1:*
192:	gb_htc:*
193:	em_gss_fun1:*
194:	em_gss_hum1:*
195:	em_gss_hum2:*
196:	em_gss_hum3:*
197:	em_gss_hum4:*
198:	em_gss_hum5:*
199:	em_gss_hum6:*
200:	em_gss_hum7:*
201:	em_gss_hum8:*
202:	em_gss_hum9:*
203:	em_gss_inv1:*
204:	em_gss_inv2:*
205:	em_gss_inv3:*
206:	em_gss_inv4:*
207:	em_gss_pln1:*
208:	em_gss_pln2:*
209:	em_gss_pro:*
210:	em_gss_rdd1:*
211:	em_gss_rdd2:*
212:	em_gss_rdd3:*
213:	em_gss_rdd4:*
214:	em_gss_rdd5:*
215:	em_gss_vrt1:*
216:	em_gss_vrt2:*
217:	em_gss_vrt3:*
218:	gb_gss1:*
219:	gb_gss2:*
220:	gb_gss3:*
221:	gb_gss4:*
222:	gb_gss5:*
223:	gb_gss6:*
224:	gb_gss7:*
225:	gb_gss8:*
226:	gb_gss9:*
227:	gb_gss10:*
228:	gb_gss11:*
229:	gb_gss12:*
230:	gb_gss13:*
231:	gb_gss14:*
232:	gb_gss15:*
233:	gb_gss16:*
234:	gb_gss17:*
235:	gb_gss18:*
236:	gb_gss19:*
237:	gb_gss20:*
238:	gb_gss21:*
239:	gb_gss22:*
240:	gb_gss23:*
241:	gb_gss24:*
242:	gb_gss25:*
243:	gb_gss26:*
244:	gb_gss27:*
245:	gb_gss28:*
246:	gb_gss29:*
247:	gb_gss30:*
248:	gb_gss31:*
249:	gb_gss32:*
250:	gb_gss33:*
251:	gb_gss34:*
252:	em_gss_inv4:*
253:	em_gss_rdd6:*
254:	em_gss_rdd7:*
255:	em_gss_rdd8:*
256:	gb_gss35:*
257:	gb_gss36:*
258:	gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	343.8	95.5	523	142	BE958470 601644787
2	302.4	84.0	304	105	AL047637 DKFZp586J
3	302	83.9	851	144	BF102859 601646827
4	283.8	78.8	950	143	BF031624 601558104
5	274.2	76.2	523	143	BF021679 uY50F05.y
6	265.2	73.7	556	155	BE609411 323251.NA
7	261.8	72.7	564	145	BF151355 uz15b12.y
8	255.8	71.1	874	168	BF670092 602119417
9	250.2	69.5	910	175	BG243001 602355974
10	230.2	63.9	447	143	BF011835 us37d10.y
11	209.2	58.1	968	174	BG118707 602348280
12	158.4	44.0	902	145	BF163629 601769732
13	141.4	39.3	425	11	AA759138 ah78f03.s
14	139.6	38.8	521	166	BE374398 601227568
15	118.6	32.9	324	15	AI020581 ua96c08.r
16	116.6	32.4	877	220	CNS024MX Tetraodon
17	116.4	32.3	337	112	AW167693 xn47g09.x
18	116.4	32.3	389	15	AI024617 ov52f10.x
19	96.4	26.8	511	113	AW210331 ul53f09.y
20	94.6	26.3	126	144	BF117246 uz07a06.y
21	89.4	24.8	354	20	AI470494 t144ell.x
22	89	24.7	1796	192	AK003359 Mus muscu
23	83.8	23.3	192	118	AW582560 RCL-ST027
24	75.6	21.0	1036	221	CNS0458A Tetraodon
25	69.4	19.3	765	145	BF143048 601788733
26	68.6	19.1	155	142	BE937452 RCL-ST027
27	54.2	15.1	649	165	BE289243 601092806
28	53	14.7	560	122	AW968905 EST380981
29	41	11.4	412	190	W21436 zb2a06.r1
30	39.6	11.0	1029	174	BG173847 602333906
31	38.8	10.8	828	221	CNS030LN Tetraodon
32	36.6	10.2	877	145	BF137533 601780532
33	36	10.0	696	153	BG430400 602502113
34	34.4	9.6	584	240	AZ243830 RPL-23-4
35	34.2	9.5	444	143	BE990576 UI-M-B21-
36	34.2	9.5	445	143	BE990947 UI-M-B21-
37	34.2	9.5	547	141	BE844607 EST352 Ap
38	34.2	9.5	1059	221	AL265760 Tetraodon
39	33.8	9.4	400	241	AZ279383 RPL-23-1
40	33.8	9.4	479	190	W76903 me58b06.r1
41	33.8	9.4	506	256	B78465 T30C13FFB.T
42	33.6	9.3	149	142	BE937398 RCL-ST027
43	33.6	9.3	336	108	AV024535 AV024535
44	33.4	9.3	357	7	AA424616 zv91c03.r
45	33.4	9.3	481	148	BF433302 7q60f09.x

ALIGNMENTS

Result No.	Score	Query Match %	Length	DB ID	Description
1	343.8	95.5	523	142	BE958470 601644787
2	302.4	84.0	304	105	AL047637 DKFZp586J
3	302	83.9	851	144	BF102859 601646827
4	283.8	78.8	950	143	BF031624 601558104
5	274.2	76.2	523	143	BF021679 uY50F05.y
6	265.2	73.7	556	155	BE609411 323251.NA
7	261.8	72.7	564	145	BF151355 uz15b12.y
8	255.8	71.1	874	168	BF670092 602119417
9	250.2	69.5	910	175	BG243001 602355974
10	230.2	63.9	447	143	BF011835 us37d10.y
11	209.2	58.1	968	174	BG118707 602348280
12	158.4	44.0	902	145	BF163629 601769732
13	141.4	39.3	425	11	AA759138 ah78f03.s
14	139.6	38.8	521	166	BE374398 601227568
15	118.6	32.9	324	15	AI020581 ua96c08.r
16	116.6	32.4	877	220	CNS024MX Tetraodon
17	116.4	32.3	337	112	AW167693 xn47g09.x
18	116.4	32.3	389	15	AI024617 ov52f10.x
19	96.4	26.8	511	113	AW210331 ul53f09.y
20	94.6	26.3	126	144	BF117246 uz07a06.y
21	89.4	24.8	354	20	AI470494 t144ell.x
22	89	24.7	1796	192	AK003359 Mus muscu
23	83.8	23.3	192	118	AW582560 RCL-ST027
24	75.6	21.0	1036	221	CNS0458A Tetraodon
25	69.4	19.3	765	145	BF143048 601788733
26	68.6	19.1	155	142	BE937452 RCL-ST027
27	54.2	15.1	649	165	BE289243 601092806
28	53	14.7	560	122	AW968905 EST380981
29	41	11.4	412	190	W21436 zb2a06.r1
30	39.6	11.0	1029	174	BG173847 602333906
31	38.8	10.8	828	221	CNS030LN Tetraodon
32	36.6	10.2	877	145	BF137533 601780532
33	36	10.0	696	153	BG430400 602502113
34	34.4	9.6	584	240	AZ243830 RPL-23-4
35	34.2	9.5	444	143	BE990576 UI-M-B21-
36	34.2	9.5	445	143	BE990947 UI-M-B21-
37	34.2	9.5	547	141	BE844607 EST352 Ap
38	34.2	9.5	1059	221	AL265760 Tetraodon
39	33.8	9.4	400	241	AZ279383 RPL-23-1
40	33.8	9.4	479	190	W76903 me58b06.r1
41	33.8	9.4	506	256	B78465 T30C13FFB.T
42	33.6	9.3	149	142	BE937398 RCL-ST027
43	33.6	9.3	336	108	AV024535 AV024535
44	33.4	9.3	357	7	AA424616 zv91c03.r
45	33.4	9.3	481	148	BF433302 7q60f09.x

Result No.	Score	Query Match %	Length	DB ID	Description
1	343.8	95.5	523	142	BE958470 601644787
2	302.4	84.0	304	105	AL047637 DKFZp586J
3	302	83.9	851	144	BF102859 601646827
4	283.8	78.8	950	143	BF031624 601558104
5	274.2	76.2	523	143	BF021679 uY50F05.y
6	265.2	73.7	556	155	BE609411 323251.NA
7	261.8	72.7	564	145	BF151355 uz15b12.y
8	255.8	71.1	874	168	BF670092 602119417
9	250.2	69.5	910	175	BG243001 602355974
10	230.2	63.9	447	143	BF011835 us37d10.y
11	209.2	58.1	968	174	BG118707 602348280
12	158.4	44.0	902	145	BF163629 601769732
13	141.4	39.3	425	11	AA759138 ah78f03.s
14	139.6	38.8	521	166	BE374398 601227568
15	118.6	32.9	324	15	AI020581 ua96c08.r
16	116.6	32.4	877	220	CNS024MX Tetraodon
17	116.4	32.3	337	112	AW167693 xn47g09.x
18	116.4	32.3	389	15	AI024617 ov52f10.x
19	96.4	26.8	511	113	AW210331 ul53f09.y
20	94.6	26.3	126	144	BF117246 uz07a06.y
21	89.4	24.8	354	20	AI470494 t144ell.x
22	89	24.7	1796	192	AK003359 Mus muscu
23	83.8	23.3	192	118	AW582560 RCL-ST027
24	75.6	21.0	1036	221	CNS0458A Tetraodon
25	69.4	19.3	765	145	BF143048 601788733
26	68.6	19.1	155	142	BE937452 RCL-ST027
27	54.2	15.1	649	165	BE289243 601092806
28	53	14.7	560	122	AW968905 EST380981
29	41	11.4	412	190	W21436 zb2a06.r1
30	39.6	11.0	1029	174	BG173847 602333906
31	38.8	10.8	828	221	CNS030LN Tetraodon
32	36.6	10.2	877	145	BF137533 601780532
33	36	10.0	696	153	BG430400 602502113
34	34.4	9.6	584	240	AZ243830 RPL-23-4
35	34.2	9.5	444	143	BE990576 UI-M-B21-
36	34.2	9.5	445	143	BE990947 UI-M-B21-
37	34.2	9.5	547	141	BE844607 EST352 Ap
38	34.2	9.5	1059	221	AL265760 Tetraodon
39	33.8	9.4	400	241	AZ279383 RPL-23-1
40	33.8	9.4	479	190	W76903 me58b06.r1
41	33.8	9.4	506	256	B78465 T30C13FFB.T
42	33.6	9.3	149	142	BE937398 RCL-ST027
43	33.6	9.3	336	108	AV024535 AV024535
44	33.4	9.3	357	7	AA424616 zv91c03.r
45	33.4	9.3	481	148	BF433302 7q60f09.x

RESULT 2

Result No.	Score	Query Match %	Length	DB ID	Description
1	343.8	95.5	523	142	BE958470 601644787
2	302.4	84.0	304	105	AL047637 DKFZp586J
3	302	83.9	851	144	BF102859 601646827
4	283.8	78.8	950	143	BF031624 601558104
5	274.2	76.2	523	143	BF021679 uY50F05.y
6	265.2	73.7	556	155	BE609411 323251.NA
7	261.8	72.7	564	145	BF151355 uz15b12.y
8	255.8	71.1	874	168	BF670092 602119417
9	250.2	69.5	910	175	BG243001 602355974
10	230.2	63.9	447	143	BF011835 us37d10.y
11	209.2	58.1	968	174	BG118707 602348280
12	158.4	44.0	902	145	BF163629 601769732
13	141.4	39.3	425	11	AA759138 ah78f03.s
14	139.6	38.8	521	166	BE374398 601227568
15	118.6	32.9	324	15	AI020581 ua96c08.r
16	116.6	32.4	877	220	CNS024MX Tetraodon
17	116.4	32.3	337	112	AW167693 xn47g09.x
18	116.4	32.3	389	15	AI024617 ov52f10.x
19	96.4	26.8	511	113	AW210331 ul53f09.y
20	94.6	26.3	126	144	BF117246 uz07a06.y
21	89.4	24.8	354	20	AI470494 t144ell.x
22	89	24.7	1796	192	AK003359 Mus muscu
23	83.8	23.3	192	118	AW582560 RCL-ST027
24	75.6	21.0	1036	221	CNS0458A Tetraodon
25	69.4	19.3	765	145	BF143048 601788733
26	68.6	19.1	155	142	BE937452 RCL-ST027
27	54.2	15.1	649	165	BE289243 601092806
28	53	14.7	560	122	AW968905 EST380981
29	41	11.4	412	190	W21436 zb2a06.r1
30	39.6	11.0	1029	174	BG173847 602333906
31	38.8	10.8	828	221	CNS030LN Tetraodon
32	36.6	10.2	877	145	BF137533 601780532
33	36	10.0	696	153	BG430400 602502113
34	34.4	9.6	584	240	AZ243830 RPL-23-4
35	34.2	9.5	444	143	BE990576 UI-M-B21-
36	34.2	9.5	445	143	BE990947 UI-M-B21-
37	34.2	9.5	547	141	BE844607 EST352 Ap
38	34.2	9.5	1059	221	AL265760 Tetraodon
39	33.8	9.4	400	241	AZ279383 RPL-23-1
40	33.8	9.4	479	190	W76903 me58b06.r1
41	33.8	9.4	506	256	B78465 T30C13FFB.T
42	33.6	9.3	149	142	BE937398 RCL-ST027
43	33.6	9.3	336	108	AV024535 AV024535
44	33.4	9.3	357	7	AA424616 zv91c03.r
45	33.4	9.3	481	148	BF433302 7q60f09.x

RESULT 2

AL047637	AL047637	304 bp	mrna	EST	01-MAR-2000
LOCUS	DFRZP386J0421_r1	586	(synonym: hutel)	Homo sapiens	cDNA clone
DEFINITION	DFRZP386J0421		mrna sequence.		
ACCESSION	AL047637				
VERSION	AL047637				
KEYWORDS	AL047637.1	GI:4728633			
SOURCE	human.				
ORGANISM	Homo sapiens				

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 851)
TITLE	NIH-MGC http://mgc.nci.nih.gov/ , National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov
COMMENT	Tissue Procurement: DCTD/DRP cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: L1CM921 row: f column: 08 High quality sequence stop: 529.

```

FEATURES
  high quality sequence stop: 323.
  Location/Qualifiers
    1..851
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:4073095"
      /clone_lib="NIH_MGC_60"
      /tissue_type="adipocarcinoma"

```

```

/clone="IMAGE:4073095"
/clone_lib="NIH_MGC_60"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site_1: SfiI (ggccgctcgccc); Site_2: SfiI (ggccattatc)";
Double-stranded cDNA was prepared from a cell line RNase
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CAGCGCATATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTGAGCCGCGCGCGCATG-3' (30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.5 kb (range 0.9-4.0 kb). 11/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."

```

BASE COUNT 234 a 169 c 238 g 210 t

ORIGIN

BASE COUNT	234 a	169 c	238 g	210 t	
ORIGIN					
Query Match	83.9%	Score 302;	DB 144;	Length 851;	
Best Local Similarity	96.1%;	Pred. No. 1.le-83;			
Matches 342; Conservative	0;	Mismatches 10;	Indels	4;	Gaps
QOy	5	ttgttttggagaataatccagagtggtgagatctgaaccttctaacaagagggttaagat	64		
Ddb	199	TTTGTTTTGGAGAATAATCCAGAGTGGTGGATCTGAACCTCTTACAGAGGAGGTAAAGAT	258		
QOy	65	tatacagctgcacacctgcgttaacttcagttgcataagggagaagaactaaagagacacg	124		
Ddb	259	TATACAGCTGCACACCTCGTACTTCTCAGTGTCCATAGGGAGAACTAATACAGAACCG	318		
QOy	125	ataccattttctggccaggtgtgtctcctggtttaaacgctgtggtggaaactgtgcctgtt	184		
Ddb	319	ATACCAATTTCTGGCCAGGTTGTCTCCTGGTTAAACGCTGTGGTGGAACTGTGGCTGT	377		
QOy	185	gtctccacaattgcaatgcaatgtcaatgtgtccccaagcaaaagtactaaaaataaccacg	244		
Ddb	378	GTCTCCCAAAATTGCAATGAAATGTCAATGTGTCCCAAGCAAAAGTTACTAAAAAATACCACG	437		
QOy	245	aggtccttcagttagagacaaagacgcgtgtgcagggggattgcacaaatactaccacgcagc	304		
Ddb	438	AGGTCCTTCAGTTGAGACCAAGACCGGTGTCCAGGGGA--TGCACAAATCACTACCCGACG	496		

[illegible]

RESULT 6
 BG609411 556 bp mRNA EST 17-APR-2001
 LOCUS 323251 MARC IP1G Sus scrofa cDNA 5', mRNA sequence.
 DEFINITION BG609411
 ACCESSION BG609411.1 GI:13659390
 VERSION EST.
 KEYWORDS
 SOURCE
 ORGANISM
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 556)
 REFERENCE
 AUTHORS
 Fahnenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
 and Keele, J.W.
 TITLE
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 JOURNAL
 COMMENT
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt trimmed with phred
 v0.980904.e. Vector identified by cross_match with the -minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACGTCACGACG
 Plate: 98 row: p column: 8
 Seq primer: ATTAGTGACACTATAG.

FEATURES
 source
 1..556
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC IP1G"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 139 a 153 c 141 g 123 t
 ORIGIN
 Query Match 73.7%; Score 265.2; DB 155; Length 556;
 Best Local Similarity 86.0%; Pred. No. 3.4e-72;
 Matches 294; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 1 gctttgttttgaagaaatccagagtggtggtatcgaaccttctaagaggagga 60
 Db 215 GCTTTGTTTTCGGAAGAAATCCAGAGTGGTGATCTGAACCTTCTCAAGGAAGAGTG 274
 Qy 61 agattatcacgtgcacacctgtgaactctcagtgccataaggaagaaactaaagaga 120
 Db 275 AGGCTATACAGTGCACCCCTAGGAACCTTTCAGTGTCTATCAGGGAAGAGCTGAGAGA 334
 Qy 121 accgatacatttttggccagggtgtctcctcgtgtaaacgctgtggtggaactgtgcc 180
 Db 335 ACAGACACCATCTTGTGGCCAGGCTGCCTCTCGTCAAGCGTGTGTGGCGGGAAGTGGCC 394
 Qy 181 tgtgtctcccaattgaatgaatgaatgtcccaagcaaaagttaactaaaaatac 240
 Db 395 TGCTGCATGCACACTGCAATGAGTGTGAGTGTGTGTCCTCCAGCAAGTCAACGAATAAT 454
 Qy 241 cacgaggtccttcagttgagacccaagaccggtgtgtcaggggattgcacaaatcacacc 300
 Db 455 CACGAGGTCTCTTCAGTTGAGACCCACAGACAGGTGTCCGGGGCTGCACAAAGTCCCTCACC 514
 Qy 301 gacgtggccctggagaccatgagaggtgtgactgtgtgtgc 342

Db 515 GACGTGGCCCTGGAGCACACGAGGAGTGTGACTGTGCGTGTGC 556
 RESULT 7
 BF151355 564 bp mRNA EST 29-DEC-2000
 LOCUS uz15b12.y1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3669119 5'
 DEFINITION similar to TR:Q9QV71 Q9QV71 FALLOTEIN.; mRNA sequence.
 ACCESSION BF151355
 VERSION BF151355.1 GI:11032750
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 564)
 REFERENCE
 AUTHORS
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE
 Tumor Gene Index
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs@mail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/ILNL at:
 image.llnl.gov/image/html/iresources.shtml

MGI:1429887
 Seq primer: -4ORP from Gibco
 High quality sequence stop: 436.
 FEATURES
 source
 1..564
 /organism="Mus musculus"
 /strain="C57/B6"
 /db_xref="taxon:10090"
 /clone_lib="IMAGE:3669119"
 /clone_lib="NCI_CGAP_Mam5"
 /tissue_type="tumor, gross tissue"
 /dev_stage="7 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Robin Humphreys,
 NIH"

BASE COUNT 157 a 130 c 140 g 137 t
 ORIGIN
 Query Match 72.7%; Score 261.8; DB 145; Length 564;
 Best Local Similarity 83.9%; Pred. No. 3.9e-71;
 Matches 296; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 1 gctttgttttgaagaaatccagagtggtggtatcgaaccttctaagaggagga 60
 Db 211 GCTTCTGTATGGGAAAAAAGCAAGTGTGTGAATCTCTCTCAAGGAAGAGGTA 270
 Qy 61 agattatcacgtgcacacctgtgaactctcagtgccataaggaagaaactaaagaga 120
 Db 271 AAACCTACAGTGCACACCCCGGAACCTTCTCAGTGTCCATACGGGAAGAGTAAAGAGG 330
 Qy 121 accgatacatttttggccagggtgtctcctcgtgtaaacgctgtggtggaactgtgcc 180
 Db 331 ACAGATACCATATATTGGCCAGGTTGTCTCTGTCTCAAGCGCTGTGAGAAATTTGTGCC 390
 Qy 181 tgtgtctcccaattgaatgaatgaatgtcccaagcaaaagttaactaaaaatac 240
 Db 391 TGTGTCTCCATATGCAATGAATGTGACTGTGTCCACGACGTAAGATTACAAAAAGATAC 450

Ddb	189	CCGATACCATGTCTTCTGGCCAGGTGTGTCTCCTGTTAAACGCCTGTGTGGGAACACTGTGCC	248
Oy	181	tgttgtccacaattgcaatgaatgtcaa-tgtgtcccacagcaaaagttaaaaaata	239
Ddb	249	TGTGGTCTCCACAATTGCAATGTAATGTCAAGTGTCGCCAAGCAAAAGTTACTAAAAATA	308
Oy	240	cga-cgagtgcttcaggttgagaccaaaagaccgggtgcaggggattgcacacatcactca	298
Ddb	309	CCAGCGAGGTCCYTTCAGTTGAGACCACCAAGACCGGTGTGAGGGATGGCACAAATCACTCA	368
Oy	299	cgaactgcccttggagaccaccatgagagtg--actgtgtgcagagggagacag	355
Ddb	369	CCGACGTGCCCTGGAGACGCATGAGAGAGTGTGCTGTGTGCAGAGGAGCAAA	428
Oy	356	g 356	
Ddb	429	G 429	

RESULT	9
BG243001	
LOCUS	602355974F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4483938 5'
DEFINITION	mRNA sequence.
ACCESSION	BG243001
VERSION	BG243001.1 GI:12752725
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Gilbert Smith, Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10323 row: d column: 19 High quality sequence stop: 690.

FEATURES	source
i . 910	Location/Qualifiers
/organism="Mus musculus"	
/strain="FVB/N"	
/db_xref="taxon:10090"	
/clone="IMAGE:4483938"	
/clone_lib="NCI_CGAP_Mam1"	
/tissue_type="tumor, biopsy sample"	
/dev_stage="3 months, virgin"	
/lab_host="DH10B"	
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: Salt; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dh. Library constructed by Life Technologies, Investigator providing samples: Gilbert Smith, NIH"	
242 a 220 c 250 g 198 t	

BASE COUNT	Query Match	69.5%;	Score 250.2;	DB 175;	Length 910;
ORIGIN	Best Local Similarity	84.7%;	Pred. No. 1.9e-67;		
	Matches 304;	Conservative	0;	Mismatches 53;	Indels 2;
				Gaps	2;

Oy	1	gctttgttttttgaagaataatccaagtggtgatctgaacctttaacagaggaggtta	60
Ddb	155	GCTTTCTGTATGGGAAAACAAAGTGGTGAATCTGAATCTCTCAAGGAAGAGGTA	214
Oy	61	agattatacagctgcacacctgtgaacttctcaqtgtccataaaggaagaactaaagaqa	120

Db	215	AAACTCTACAGTCACACCCCGAACTTCTCAGTGTCCATACGGAAGAGCTAAAGGG	274
QY	121	accgataccatttctggccaggttctctctgtgttaaacgcgtgtgtggaactgtgcc	180
Db	275	ACAGATACCAATATCTCTGGCAGGTTCTCTCTGGTCAAGCGCTGTGG-AGGAAATGTGCC	333
QY	181	tgttctccacaattgcaatgaatgtcaatgtgtcccaagcaaaagtctactaaaaataac	240
Db	334	TGTTGTCTCCATAAATTTGCAATGAATGTCAAGTGTGTGTCCACGTAAAGTTACAAAAAGTAC	393
QY	241	cacgaggtctctcagttgagacaaagaccggtgtgcaggggattgcacaaatcactcacc	300
Db	394	CATGAGTCTCTCAGTTGAGACCAA-ACCTGAGTCAAGGGATTGCATAAGTCACCTACT	452
QY	301	gacgtggccctggagcaccatgaggagtgactgtgtgtgcagaggagcacaggagg	359
Db	453	GATGTGGCTCTGGAACACACCGAGGAATGTGACTGTGTGTGTGTAGAGGAAACCGCAGGAGG	511

RESULT 10

BF011835

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BF011835

447 bp

mRNA

EST

06-OCT-2000

us37d10.y1 Soares_NMEBA_branchial_arch Mus musculus cDNA clone

IMAGE:3169267 5' similar to TR:Q90Y71 Q90Y71 FALLOTEIN. ;, mrna

sequence.

BF011835

GI:10712110

BF011835.1

house mouse.

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 447)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo

, Ph.D.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

image.llnl.gov/image/html/iresources.shtml

MGI:1064727

Seq primer: -40RP from Gibco

High quality sequence stop: 402.

Location/Qualifiers

1. 447

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone_lib="Soares_NMEBA_branchial_arch"

/tissue_type="branchial arches"

/dev_stage="embryo, 10.5 dpc"

/lab_host="DH10B (phage resistant)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: NotI; Site_2: EcoRI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTGGAGCGCGCATGCTATTTTTTTTTTTTTTTTTTTTTTTTTT

3']; double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not

I and Eco RI sites of the modified pT73 vector. Library

constructed and-normalized by Bento Soares and M.Fatima

Bonaldo."

FEATURES

source

BASE COUNT

126 a 104 c 108 g 109 t

ORIGIN

Query Match	63.9%;	Score 230.2;	DB 143;	Length 447;					
Best Local Similarity	85.6%;	Pred. No. 2.9e-61;							
Matches 256;	Conservative	0;	Mismatches 43;	Indels 0;					
Gaps									
QY 1	gctttgttttggagaataaccagagtggttgatctgaaccttctacaagaggagga	60							
Db 148	GCYTTCTGTATGGGAAAAAACCAAGTGGTGAATCTGAATCTCTCAAGGAAGGTA	207							
QY 61	agattatcacgctgcacactcgttaacttctcagtgctccataagggaagaaptaa	120							
Db 208	AAACTACAGCTGCACACCCCGAACTTCTCAGTGTCCATACGGGAAGAGTAA	267							
QY 121	acgatacattcttctggcaggttctctctcgttggttaaacgctgtgtgtgga	180							
Db 268	ACAGATACCATATATCTGGCAGGTTGTCTCTGGTCAAGCGCTGTGGAGGAAT	327							
QY 181	tggtgtccacaattgcaataatctcaatgtgtcccaagcaaatgtactaaaaa	240							
Db 328	TGTTGTCTCCATTAATGTCAATGAATGTCTAGTGTGCCAGTAAAGTTACA	387							
QY 241	cacgaggtctctcagttgaccaaagaccggtgtcaggggattgcacaaacac	299							
Db 388	CATGAGGTCTCTCAGTTGAGACCAAACTGGAGTCAAGGGATTGATAAATCA	446							
RESULT 11									
BGI18707	968 bp	mRNA	EST	30-JAN-2001					
LOCUS	602348280F1	NIH_MGC_90	Homo sapiens	cdna clone IMAGE:4443498 5'					
DEFINITION	mRNA sequence.								
ACCESSION	BGI18707								
VERSION	BGI18707.1	GI:12612213							
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
TITLE	1 (bases 1 to 968)								
JOURNAL	NIH-MGC	http://mgc.nci.nih.gov/.							
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)								
	Unpublished (1999)								
	Contact: Robert Strausberg, Ph.D.								
	Email: cgapbs-remail.nih.gov								
	Tissue Procurement: ATCC								
	CDNA Library Preparation: Life Technologies, Inc.								
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)								
	DNA Sequencing by: Incyte Genomics, Inc.								
	Clone distribution: MGC clone distribution information can be								
	found through the I.M.A.G.E. Consortium/LNL at:								
	http://image.llnl.gov								
	Plate: LLAM10217 row: 0 column: 19								
	High quality sequence start: 8								
	High quality sequence stop: 574.								
FEATURES	Location/Qualifiers								
source	1..968								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/clone="IMAGE:4443498"								
	/clone_lib="NIH_MGC_90"								
	/tissue_type="adenocarcinoma, cell line"								
	/lab_host="DH10B (phage-resistant)"								
	/note="Organ: liver; Vector: pCMV-Sport6; Site:1: NotI;								
	Site:2: SalI; Cloned unidirectionally; oligo-dT primed.								
	Average insert size 1.7 kb. Library enriched for								
	full-length clones and constructed by Life Technologies.								
	Note: this is a NIH_MGC Library."								
BASE COUNT	313 a	230 c	219 g	206 t					
ORIGIN									
Query Match	58.1%;	Score 209.2;	DB 174;	Length 968;					
Best Local Similarity	94.2%;	Pred. No. 1.3e-54;							
Matches 228;	Conservative	0;	Mismatches 13;	Indels 1;					
Gaps									

DB	13	CGCTGGGAGGAAATTGTCCTCTTCTCCATAAATGCAATGTCAGTGTGTCCTCA	72
QY	220	agcaagttactaaaaataaccacgaggtccttcagttgagacccaagaccgtgtcagg	279
Db	73	CGTAAAGTTACAAAAAGTACCATGAGGTCCCTTCAGTTGAGACCAAAAACTGAGTCAAG	132
QY	280	ggattgcacaaatcactaccacgagctggccctggagacaccatgagagtggtgactgtgtg	339
Db	133	GGATTGCATTAAGTCACTCACTGATGTGGTCTGGAACACACCCAGGAAATGACTGTGTG	192
QY	340	tgacaggggagcacaggagg	359
Db	193	TGTAGAGGAAACCGCAGGAG	212

RESULT	13
AA759138/c	
LOCUS	AA759138 425 bp mRNA EST 29-DEC-1998
DEFINITION	ah78f03.s1 Soares_testis_NHT Homo sapiens cDNA clone 1321757 3', mRNA sequence.
ACCESSION	AA759138
VERSION	AA759138.1 GI:2807001
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 425)
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
FEATURES	cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bbrp/image/image.html Insert Length: 1417 Std Error: 0.00 Seq primer: -40m13 fwd. Et from Amersham High quality sequence stop: 385. Location/Qualifiers 1. 425 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="1321757" /clone_lib="Soares_testis_NHT" /sex="male" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer (5' TGTTACCAATCTGAAGTGGAGCGCCGCCAATTTTTTTTTTTT 3'). Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT	100 a 108 c 87 g 130 t
ORIGIN	
Query Match	39.3%; Score 141.4; DB 11; Length 425;
Best Local Similarity	98.7%; Pred. No. 1.8e-33;
Matches	153; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	206	gtcaatgtgtcccaagcaaaagtactataaaaaataccacaggtctctcagttgagaccaa	265
Db	425	GTCAATGTGTCCCAAGCAAAAGTACTATAAAAAATACCACGAGTCTCTCAGTTGAGACCAA	366

```
Db      183 ACCGAGGG 192
```

```
RESULT 15  
AI020581  
LOCUS   Soaresmammary_gland_NBMGM Mus musculus cDNA clone  
DEFINITION  
ACCESSION U96C08 rl Soaresmammary_gland_NBMGM EST       16-JUN-1998  
IMAGE:1365326 5' mRNA sequence.  
AI020581  
AI020581.1 GI:3234917  
EST.  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 324)  
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubouque,T.,  
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilsson,R. and  
Waterston,K.
```

```
TITLE The WashU-HHMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mousee@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMGE Consortium (info@image.llnl.gov) for further information.  
MG1:898546  
Seq primer: ~28ml3 rev2 ET from Amersham  
High quality sequence stop: 283.  
Location/Qualifiers  
1..324  
    organism="Mus musculus"  
    strain="CS7BL/6J"  
    db_xref="taxon:10090"  
    clone="IMAGE:1365326"  
    clone_lib="Soares_mammary_gland_NBMMG"  
    sex="male"  
    tissue_type="mammary gland"  
    dev_stage="4 weeks"  
    lab_host="DH10B"
```

```
/note=Organ: mammary gland; Vector: pT73D-Pac (Pharmacia )  
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;  
R1; lst strand cdna was primed with a Not I - oligo(dT) primer [5,  
TGTTACCATCTGAAGTCGGAGCGGCCGGTAATGGTTTTTTTTTTTTTTTTTTTTTTT  
T 3']; double-stranded CDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I and  
Eco RI sites of the modified pT73 vector. RNA provided by Dr.  
Minoru Ko, Wayne State Univ. Library constructed and normalized  
by Bento Soares and M.Fatima Bonaldo."
```

```
BASE COUNT      82 a    81 g     80 t  
ORIGIN  
Query Match          32.9%; Score 118.6; DB 15; Length 324;  
Best Local Similarity 82.4%; Pred.No. 2.3e-26;  
Matches 136; Conservative 0; Mismatch 29; Indels 0; Gaps 0;
```

```
OY 1 gcctttgttttggaagaaattccagagtgggtggtcctaagaaccttaacaggaggta 60  
                |||||         ||| | ||| | ||| | ||| | ||| | ||| |  
DB 160 GCTTCCGTGTGTGNAAAAGAACAAGTGGTGATCTCAATTCTCCTCAAGSAAAGGTA 219  
                |||||         ||| | ||| | ||| | ||| | ||| | ||| |  
OY 61 agattatacgctgcacacttcgatctcacttcagtgtcccataaggggaagaaactaaaga 120  
                |||||         ||| | ||| | ||| | ||| | ||| | ||| |  
DB 220 AAACTCTACAGCTGCACACCCGGGAACCTCTCACGTCTCCATACGGGAACAGCTAAGAGG 279  
                |||||         ||| | ||| | ||| | ||| | ||| | ||| |
```

Qy 121 accgataccattttcttgccaggtgtctctcctggttaaacgctgt 165
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 ACAGATACCATAATCTGGCCAGGTGTCTCCTGTGTCACAGGCTGT 324

Search completed: July 18, 2001, 03:49:55
Job time: 13553 sec

7